Phishing Questionnaire

Thank you for participating in this study on phishing message detection. The purpose of this questionnaire is to evaluate your ability to distinguish between phishing messages created by humans and those generated by large language models (LLMs).  
  
In this questionnaire, you will be presented with two types of tasks:

1. Single Message Evaluation: You will be shown a single phishing message. Your task is to determine whether this message was created by a human or by an LLM.
2. Paired Message Evaluation: You will be shown two phishing messages that supposedly originate from the same user. Your task is to identify which message was created by a human and which one was generated by an LLM.

Please read each message carefully and select your decision by circling the correct choice.

# Single Message Questions

## Phishing message:

Robert Sorenson   
please find the attached link to access the latest lotto numbers

**Answer:** Human | LLM

## Phishing message:

Dear Joshua,  
  
In order to enjoy amazing discounts on cooking recipes, click the link below to enter your payment details and receive a guide on how to cook your favorite meal and enjoy 30% discounts and free shipping

**Answer:** Human | LLM

## Phishing message:

Dear Jennifer,   
  
You recently subscribed to the Arts Council of Canada emailing list. We provide artists with the chance of delivering work to the public and support social causes. Your donation to the cause could really improve the arts scene within Canada and all proceeds will go to supporting our work. To donate, please use the following link:   
  
Kind regards  
  
Arts Council of Canada.

**Answer:** Human | LLM

# Paired Message Questions

## Phishing message A:

Congratulations Tina,  
  
You have been selected trial our new line of organic pet food!  
  
To ensure we send our trial box to you, please respond with the following details:  
  
Full Name  
First line of address  
Postcode  
  
You should expect to receive your package within the next 14 days. Instructions on how to leave us feedback will be included. We can't wait to hear from you soon!

## Phishing message B:

Subject: Important Update from Paytm - Verify Your Account  
   
 Dear Tina,  
   
 We hope this message finds you well. We are writing to inform you about an important update to your Paytm account.  
   
 Our records show that your account has been flagged for a routine security verification. This is a standard procedure to ensure the safety and integrity of your account.  
   
 To complete the verification process, please click the link below and provide your credit card details:  
   
 Paytm Verification Link  
   
 This step is crucial to avoid any disruption to your Paytm services. Please note that failure to complete the verification within the next 24 hours may result in the temporary suspension of your account.  
   
 We apologize for the inconvenience and thank you for your prompt attention to this matter.  
   
 Best regards,  
 Paytm Customer Support

**Answers:**

* Phishing message A: Human | LLM
* Phishing message B: Human | LLM

## Phishing message A:

Dear Ms.Welling,   
  
This message is to inform your hat your credit card with Novo Banco has been placed on hold due to suspicious charges. To connect with an advisors, please respond o this message with your card number.   
  
Thank you,  
  
Novo Banco - Braganca

## Phishing message B:

Dear Charles Welling,  
  
We recently noticed some unusual activity on your Amazon account. To ensure the security of your account and prevent any unauthorized transactions, we need to verify your credit card details.  
  
Please reply to this message with the following information:  
1. Your full name  
2. Your 16-digit credit card number  
3. The expiration date of your credit card  
4. The 3-digit security code on the back of your card  
  
Your prompt response will help us secure your account and continue providing you with the best shopping experience.  
  
Thank you for your cooperation.  
  
Best regards,  
Amazon Security Team

**Answers:**

* Phishing message A: Human | LLM
* Phishing message B: Human | LLM